PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-101223
(43)Date of publication of application : 13.04.2001

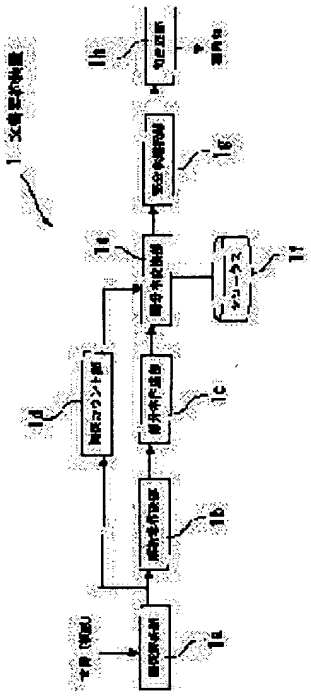
(51)Int.Cl. G06F 17/30

(21)Application number : 11-281535
(22)Date of filing : 01.10.1999

(71)Applicant : FUJI XEROX CO LTD
(72)Inventor : KOYAMA TAKEHIRO
UEDA YOSHIHIRO

(54) DOCUMENT SUMMARIZING DEVICE, DOCUMENT SUMMARIZING METHOD AND RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To create a summary from plural documents the words of which are not unified in short time.
SOLUTION: Binomial relation is extracted from texts of the plural documents 2 by a relation analyzing part 1a, the binomial relation is counted by a relation counting part 1d, a parse tree is generated by a parse tree generating part 1b, a partial tree is generated by a partial tree generating part 1c, a new partial tree is generated by converting the partial tree into the one related to the binomial relation counted by the relation counting part 1d by a partial tree converting part 1e, scores are imparted to the partial tree, the partial tree to which higher score is imparted is transmitted to a phrase synthesizing part 1h as an important partial tree by a partial tree selecting part 1g and the summary is synthesized from the transmitted important partial tree by the phrase synthesizing part 1h.



LEGAL STATUS

[Date of request for examination] 29.05.2003

[Date of sending the examiner's decision of

h g e e h

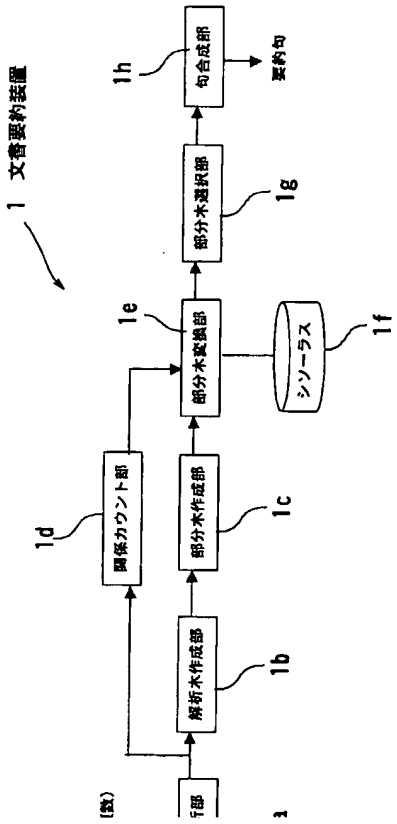
* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

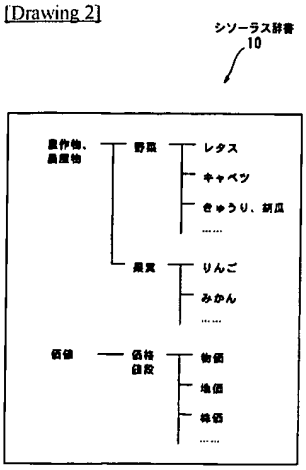
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DRAWINGS

[Drawing 1]



[Drawing 2]



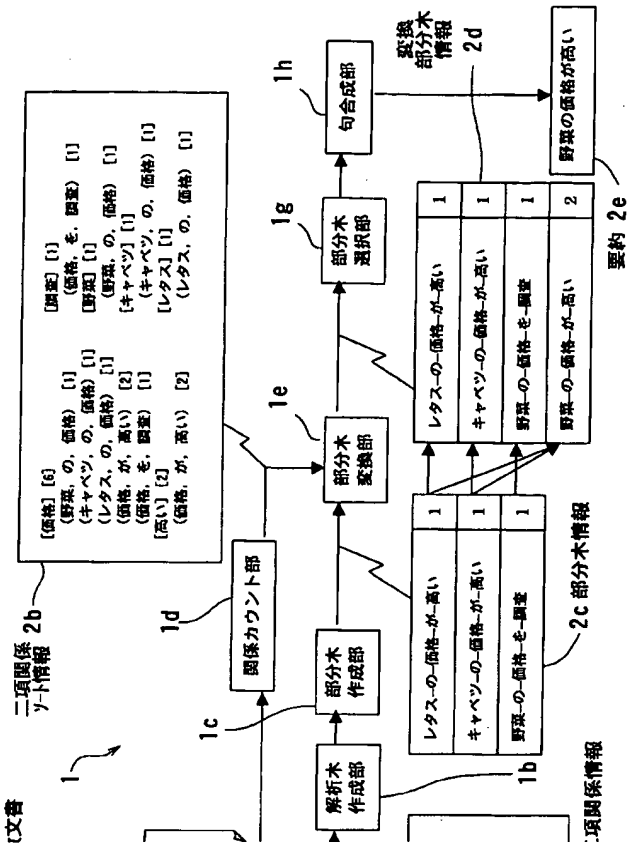
[Drawing 4]

部分木番号	部分木	頻度	スコア
1	レタス - の - 価格 - が - 高い	1	1
2	キャベツ - の - 価格 - が - 高い	1	1
3	野菜 - の - 価格 - を - 調査	1	1
4	野菜 - の - 価格 - が - 高い	2	2

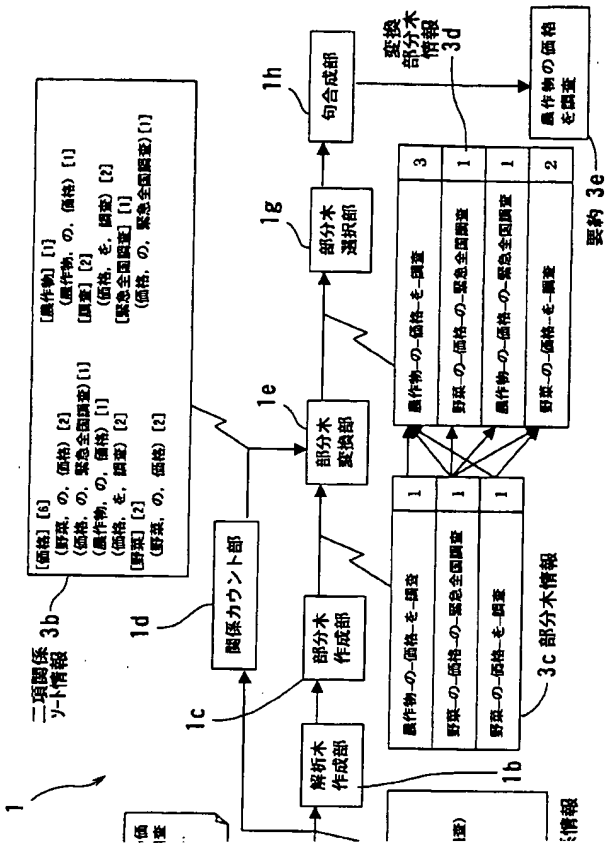
[Drawing 6]

部分木番号	部分木	頻度	スコア
1	農作物 - の - 価格 - を - 調査	3	2.19
2	野菜 - の - 価格 - の - 緊急全国調査	1	1
3	農作物 - の - 価格 - の - 緊急全国調査	1	0.7
4	野菜 - の - 価格 - を - 調査	2	1.7

[Drawing 3]


文書
1
二項関係ソート情報 2b
1d
関係カウント部
1c
部分木作成部
解析木作成部
1b
二項関係情報
1e
部分木変換部
1g
部分木選択部
1h
句合成部
最終部分木情報 2d
2c 部分木情報
レタス-の-価格-が-高い
キャベツ-の-価格-が-高い
野菜-の-価格-を-調査
野菜-の-価格-が-高い
要約 2e
野菜の価格が高い

[Drawing 5]



[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram having shown the configuration of document epitome equipment.
[Drawing 2] It is drawing having shown the contents of the thesaurus dictionary.
[Drawing 3] It is drawing having shown the epitome creation process of document epitome equipment in which two or more sentence document was inputted.
[Drawing 4] It is drawing having shown the set of the subtree to which the score was given.
[Drawing 5] It is drawing having shown the epitome creation process of document epitome equipment in which two or more sentence document was inputted.
[Drawing 6] It is drawing having shown the score of each subtree at the time of setting a score lapse rate as 0.7.
[Description of Notations]
1 Document Epitome Equipment
1a Related analysis section
1b Parse tree creation section
1c Subtree creation section
1d Related count area
1e Subtree transducer
1f Thesaurus
1g Subtree selection section
1h Phrase composition section
10 Thesaurus Dictionary

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] The program to which the document epitome equipment which creates an epitome, the document epitome approach, and its function make carry out to a computer is related in this invention to the record medium which recorded by compounding the subtree which extracted from two or more sentence document especially about the record medium which recorded the program which makes the function which creates an epitome from the document epitome equipment which creates an epitome from two or more sentence document, the document epitome approach, and two or more sentence document perform to a computer.
[0002]
[Description of the Prior Art] The various document epitome approaches which create the epitome of the contents described by two or more documents have so far been proposed. As such a document epitome approach, keyword listing (Scatter/Gather JP,5-225256,A), composition (JP,10-134066,A) of two or more sentences, etc. are typical. It is the approach of enumerating the frequent appearance words which keyword listing classifies a document group and appear in the classified cluster in the cluster, and composition of two or more sentences is the approach of creating an epitome by deleting those redundancy parts and compounding a series of document groups.
[0003] However, since it becomes enumeration of a mere word, the result which is extracted in keyword listing has the trouble that relation between words cannot be grasped. Moreover, in composition of two or more sentences, there is a problem that the field and format of an application document are limited.
[0004] As the document epitome approach which solves such a problem, an applicant for this patent With the value of the score of the subtree which was accumulated having carried out the dependency analysis of the document and carrying out score addition of the subtree which created the parse tree, and cut down and cut down subtree from the created parse tree, and was accumulated It applied for two or more sentence document epitome equipment which chooses important subtree and creates an epitome phrase from the selected important subtree (Japanese Patent Application No. 10-358596).
[0005] Generally, the word used for two or more sentence document is not unified between documents in many cases. So, with two or more sentence document epitome equipment of Japanese Patent Application No. 10-358596, the word used for two or more sentence document is synonym-developed and compound developed, subtree created by that cause is counted, and subtree with high frequency used for epitome creation is selected. For example, in the subtree (national [- urgent] investigation of - price - of lettuce -), "lettuce" is developed by the broader term "vegetables, agricultural products, and agricultural products", a "price" is developed by the synonym "a price", and "national [urgent] investigation" is developed by "national investigation and investigation."
[0006]
[Problem(s) to be Solved by the Invention] However, with two or more sentence document epitome equipment of Japanese Patent Application No. 10-358596, two or more subtrees are cut down from two or more one sentence document, and further, since synonym expansion and compound expansion will

develop at two or more subtrees, respectively, the combination of the subtree developed increases the cut-down subtree explosively. For the reason, the computational complexity and capacity which process it will increase, and there is a trouble that epitome creation will take time amount.
[0007] This invention is made in view of such a point, and it aims at offering the document epitome equipment which can create an epitome for a short time from two or more sentence document into which the word is not unified.
[0008] Moreover, other purposes of this invention are offering the document epitome approach which can create an epitome for a short time from two or more sentence document into which the word's is not unified. Furthermore, other purposes of this invention are offering the record medium which recorded the program which makes the function which creates an epitome for a short time perform to a computer from two or more sentence document into which the word's is not unified.
[0009]
[Means for Solving the Problem] In the document epitome equipment which creates the epitome of two or more inputted sentence document, and performs the output in order to solve the above-mentioned technical problem in this invention A related analysis means to perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and to extract the binary relation of words, A related count means to count the frequency of said binary relation which said related analysis means extracted, A parse tree creation means to create a parse tree from said binary relation which said related analysis means extracted, A subtree creation means to create subtree from said parse tree which said parse tree creation means created, The subtree conversion means which changes into the subtree relevant to said binary relation among which said related count means counted said subtree which said subtree creation means created, and creates new subtree, A score is given to said subtree and the document epitome equipment characterized by having a subtree selection means by which said score chooses said high subtree as important subtree, and a phrase composition means to compound an epitome by said important subtree which said subtree selection means chose is offered.
[0010] A related analysis means performs the dependency analysis of the text which extracted the text and was extracted from two or more inputted sentence document, and extracts the binary relation of words here. A related count means The frequency of the binary relation which the related analysis means extracted is counted. A parse tree creation means A parse tree is created from the binary relation which the related analysis means extracted. A subtree creation means Subtree is created from the parse tree which the parse tree creation means created. A subtree conversion means It changes into the subtree relevant to said binary relation among which said related count means counted said subtree which said subtree creation means created, and new subtree is created. A subtree selection means A score is given to subtree, a score chooses high subtree as important subtree, and a phrase composition means compounds an epitome by the important subtree which the subtree selection means chose.
[0011] Moreover, create the epitome of two or more inputted sentence document, and it sets to the document epitome approach of performing the output. Perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and the binary relation of words is extracted. Count the frequency of said binary relation and a parse tree is created from said binary relation. Create subtree from said parse tree, and change into the subtree relevant to said binary relation which counted said subtree, and new subtree is created. A score is given to said subtree, said score chooses said high subtree as important subtree, and the document epitome approach characterized by compounding an epitome by said important subtree is offered.
[0012] By doing in this way, it becomes possible to narrow down the number of the new subtrees to create, and epitome creation time can be shortened. Furthermore, perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and the binary relation of words is extracted. Count the frequency of said binary relation and a parse tree is created from said binary relation. Create subtree from said parse tree, and change into the subtree relevant to said binary relation which counted said subtree, and new subtree is created. A score is given to said subtree, said score chooses said high subtree as important subtree, and the record medium which recorded the program to which the function which compounds an epitome by said important subtree is

made to carry out to a computer is offered.
[0013] By doing in this way, it becomes possible to narrow down the number of the new subtrees to create, and epitome creation time can be shortened.
[0014]
[Embodiment of the Invention] Hereafter, the gestalt of operation of this invention is explained with reference to a drawing. First, the gestalt of the 1st operation in this invention is explained.
[0015] Drawing 1 is the block diagram having shown the configuration of the document epitome equipment 1 in this gestalt. From two or more inputted sentence document, document epitome equipment 1 extracts a text and performs the dependency analysis of the extracted text. Related analysis section 1a which extracts the binary relation of words and which is a related analysis means, 1d of related count areas which are a related count means to count the frequency of the binary relation which related analysis section 1a extracted, Parse tree creation section 1b which is a parse tree creation means to create a parse tree from the binary relation which related analysis section 1a extracted, Subtree creation section 1c which is a subtree creation means to create subtree from the parse tree which parse tree creation section 1b created, The subtree which subtree creation section 1c created is changed into the subtree relevant to the binary relation which 1d of related count areas counted. Subtree transducer 1e which is the subtree conversion means which creates new subtree, Thesaurus 1f which is the word relation record means which recorded the high order low order relation of a word, A score is given to subtree and it is constituted by 1g of subtree selection sections which are a subtree selection means by which a score chooses high subtree as important subtree, and 1h of phrase composition sections which are a phrase composition means to compound an epitome by the important subtree which 1g of subtree selection sections chose.
[0016] Drawing 2 is drawing having shown the contents of the thesaurus dictionary 10 in which the vertical relation of the word recorded on thesaurus 1f is shown. Two or more words are recorded on the thesaurus dictionary 10 with the vertical relation between each word, and the broader term, twin language, a synonym, etc. can be extracted now by choosing one word. For example, the broader terms of a word "a cucumber" are a word "vegetables" and "agricultural products and agricultural products", twin language turns into a word "lettuce", a "cabbage", etc., and a synonym turns into a word "a cucumber."
[0017] Next, actuation of document epitome equipment 1 is explained using the actual example of a document. Drawing 3 is drawing having shown the epitome creation process of document epitome equipment 1 in which two or more sentence document 2 was inputted. Here, although two or more actual sentence document 2 will consist of two or more sentences, in order to simplify explanation, it explains paying attention to one sentence each of the price of lettuce is (high), (the price of a cabbage being high), and (investigating the prices of vegetables).
[0018] First, if two or more sentence document 2 is inputted, related analysis section 1a will perform the dependency analysis of the text which extracted the text and was extracted from two or more sentence document 2, and will extract the binary relation of the words expressed with (the relation between the words by the side of charge, and the word by the side of a receptacle). the binary relation which the text of two or more sentence document 2 is the price of lettuce being (high), (the price of a cabbage being high), and (investigating the prices of vegetables) here, and is extracted -- being [of lettuce **] a (price), (price ** being high), and -- (price of vegetable **) -- it becomes investigation) about (price.
[0019] It means that investigation) has required the "price" for "investigation" by relation "**" in (price here. In addition, although the relation between words is expressed by surface relation "**" with this gestalt, with it, it is good also as expressing with the relation "an object" of the depths.
[0020] Thus, the set of the binary relation extracted by related analysis section 1a is sent to 1d of related count areas, and parse tree creation section 1b as binary relation information 2a. 1d of related count areas which received binary relation information 2a (price of lettuce **) ((price ** is high) and -- (price of vegetable **) -- (price -- investigation)) Binary relation sort information 2b in which or (word by the side of charge) (word by the side of a receptacle) constitutes binary relation, and one of words is common in the binary relation of the word which received binary relation information 2a has and which

sorted for every group and counted two term each-related frequency is created. Here, [ ] shown in binary relation sort information 2b shows frequency information. Thus, created binary relation sort information 2b is sent to subtree transducer 1e.

[0021] on the other hand, parse tree creation section 1b which received related analysis section 1a to binary relation information 2a (price of lettuce **) ((price ** is high) and -- (price of vegetable **) -- (price -- investigation)) creates a parse tree from received binary relation information 2a similarly. It becomes - investigation) about - price [of vegetable -] -. the parse tree which is created in the case of this example -- (- price [of lettuce -] - - it is high --) -- (- price [of cabbage -] - - it is high --) -- (-- here a "parse tree (- price [of lettuce -] - - high) -- "lettuce" -- "a price" -- relation "**" -- starting -- "a price" -- it is high" -- relation -- "-- it means having started by ". Thus, the created parse tree is sent to subtree creation section 1c.

[0022] Subtree creation section 1c which received the parse tree creates subtree from the received parse tree, and counts the frequency. That is, when the subtree which creates the subtree, is going to set the frequency to 1 and is going to create it when the subtree which it is going to create is not created yet is already creation ending, the frequency of the corresponding subtree [finishing / creation] is added. Although the subtree of two or more works is accomplished as a subtree of a parse tree, in order to simplify here, it makes the parse tree itself subtree. Thus, the set of the created subtree sets to subtree information 2c (- price [of lettuce -] - - high : 1 and - price [of cabbage -] - - high : 1 and - price [of vegetable -] - - investigation : 1), and is sent to subtree transducer 1e. Here, the figure after ":" expresses frequency.

[0023] If subtree information 2c sent from binary relation sort information 2b sent from 1d of related count areas and subtree creation section 1c is received, subtree transducer 1e will change received subtree, will create new subtree, and will count the frequency. Here, compound conversion and synonym conversion are used for conversion of subtree. Compound conversion is the approach of changing a compound into the substring of a right side match, and when generating new subtree by this approach, the compound which subtree has is transposed to the right side match substring of that compound. For example, if the vocabulary "document epitome equipment" which is a compound is taken for an example, the right side match substring to this will serve as "epitome equipment" and "equipment", and generation of the new subtree by this approach will be performed by transposing the vocabulary "document epitome equipment" which subtree has to "epitome equipment" and "equipment." Moreover, synonym conversion is the approach of changing a word into a synonym or a broader term, and when generating new subtree by this approach, each vocabulary which subtree has is transposed to a synonym or a broader term. For example, if a word "a cucumber" is taken for an example, in the thesaurus dictionary 10 which thesaurus 1f has, the synonym of a word "a cucumber" will serve as a "cucumber" and a broader term will serve as "vegetables" and "agricultural products and agricultural products." In this example, since the vocabulary which has a compound does not exist in the subtree shown in subtree information 2c, conversion of subtree will be performed only by synonym conversion. The procedure of the subtree conversion which subtree transducer 1e performs to below is shown.

(1) Choose one binary relation of subtree.

(2) Choose one word which constitutes the extracted binary relation.

The binary relation containing the word chosen by (3) and (2) is chosen from binary relation sort information 2b.

(4) Compare words (the word "B") other than the word chosen by (2) among words other than the word chosen by (2) among the words which constitute the binary relation chosen by (3) (the word "A"), and the word which constitutes the binary relation chosen by (1) with reference to the thesaurus dictionary 10.

[0024] Conditions (a): When the word "A" is a synonym of the word "B", create the subtree which transposed the word "B" of the binary relation chosen by (1) to the word "A" as new subtree.

Conditions (b): When the word "A" is a broader term of the word "B", create the subtree which transposed the word "B" of the binary relation chosen by (1) to the word "A" as new subtree.

[0025] Conditions (c): When the word "A" is twin language of the word "B", create the subtree which

transposed the word "B" of the binary relation chosen by (1) to the word of one high order as new subtree.
[0026] Next, the procedure of such subtree conversion is explained using actual subtree information 2c (- price [of lettuce -] - - high - price [of cabbage -] - - high - price [of vegetable -] - - investigation).
[0027] first, a word "a price" is chosen from the binary relation which chose binary relation (price of lettuce **) and was chosen from subtree (- price [of lettuce -] - - high) in (2) in (1). Next, the binary relation which contains a word "a price" in (3) is chosen from binary relation sort information 2b. Thereby, investigation) is chosen in the binary relation (price of vegetable **) which binary relation sort information 2b has, and a (price of cabbage **) (price of lettuce **) (high [price **]) (price. Next, in (4), the comparison with words (the word "B") other than the word chosen by (2) among words other than the word chosen by (2) among the words which constitute selection ********** from (3) (the word "A"), and the word which constitutes the binary relation chosen by (1) is performed. If the word "A" in the binary relation (price of vegetable **) chosen by (3) serves as "vegetables" here, the word "B" of the binary relation (price of lettuce **) chosen by (1) serves as "lettuce" and the thesaurus dictionary 10 is referred to since the "vegetables" which is the words "A" is the broader terms of the "lettuce" which is the word "B", conditions (b) are suited and the new subtree (- price [of vegetable -] - - high) which transposed "lettuce" to "vegetables" is created. about the subtree of others which are extracted from subtree (- price [of lettuce -] - - high), it does not correspond to conditions.
[0028] moreover, although the new subtree (- price [of vegetable -] - - high) which changed the word "a cabbage" of this subtree into the "vegetables" which is that broader term from subtree (- price [of cabbage -] - - high) similarly will be created since the same subtree (- price [of vegetable -] - - high) generated from subtree (- price [of lettuce -] - - high) already exists, only the frequency is added.
[0029] Moreover, since subtree (it is - investigation about - price [of vegetable -] -) does not suit conditions, subtree new from subtree (it is - investigation about - price [of vegetable -] -) is not generated. as mentioned above, 2d (- investigation: - price [of lettuce -] - - high : 1 and - price [of cabbage -] - - high : 1 and - price [of vegetable -] - 1 and - price [of vegetable -] - - high : 2) of conversion subtree information which is the set of the subtree which added the newly generated subtree (- price [of vegetable -] - - high) is created. Thus, 2d of created conversion subtree information is sent to 1g of subtree selection sections.
[0030] A score will be given and 1g of subtree selection sections will choose important subtree as each subtree with the given score, if 2d (- investigation: - price [of lettuce -] - - high : 1 and - price [of cabbage -] - - high : 1 and - price [of vegetable -] - 1 and - price [of vegetable -] - - high : 2) of conversion subtree information is acquired. Although various ways to give a score are considered, in order to simplify here, let frequency be a score. The set of the subtree by which the score was given to drawing 4 in this way is shown.
[0031] consequently, subtree (- price [of vegetable -] - - high) with the most expensive score is chosen as important subtree, and the selected important subtree is sent to 1h of phrase composition sections. if important subtree (- price [of vegetable -] - - high) is obtained, 1h of phrase composition sections will compound epitome 2e (the prices of vegetables are high), and they will perform the display.
[0032] Thus, with this gestalt, related analysis section 1a extracts binary relation from the text of two or more sentence document 2. Binary relation is counted by 1d of related count areas, a parse tree is created by parse tree creation section 1b, and subtree is created by subtree creation section 1c. By subtree transducer 1e It changes into the subtree relevant to the binary relation which counted subtree by 1d of related count areas, and new subtree is generated. By 1g of subtree selection sections Subtree to which the score was given to subtree and the high score was given is made into important subtree. In 1h of phrase composition sections delivery and 1h of phrase composition sections In order to compound an epitome from the sent important subtree, the subtree created can be narrowed down only to an important thing and it becomes possible as a result to generate an epitome phrase at a high speed by small capacity.

[0033] Next, the gestalt of the 2nd operation in this invention is explained. In the document epitome equipment 1 in the gestalt of the 1st operation, subtree transducer 1e changes subtree by compound

expansion, and this gestalt illustrates the case where 1g of subtree selection sections multiplies the frequency of subtree by the score lapse rate, and they give the score of subtree.
[0034] Drawing 5 is drawing having shown the epitome creation process of document epitome equipment 1 in which two or more sentence document 3 was inputted. Here, although two or more actual sentence document 3 will consist of two or more sentences, in order to simplify explanation, it explains paying attention to one sentence each of (the prices of agricultural products are investigated), (national [urgent] investigation of the prices of vegetables), and (investigating the prices of vegetables). In addition, since actuation of related analysis section 1a, 1d of related count areas, parse tree creation section 1b, and subtree creation section 1c is the same as that of the gestalt of the 1st operation, it omits explanation.
[0035] Binary relation sort information 3b sent from 1d of related count areas, And subtree transducer 1e which received subtree information 3c (1, national [- urgent] investigation of - price - of vegetable -: - price [of agricultural-products -] - - investigation : 1 and - price [of vegetable -] - - investigation : 1) which is the set of the subtree sent from subtree creation section 1c Received subtree is changed, new subtree is created and the frequency is counted. Here, compound conversion and synonym conversion are used for conversion of subtree.
[0036] The binary relation (national [urgent] investigation of price **) which subtree (national [- urgent] investigation of - price - of vegetable -) has has the compound "national [urgent] investigation", and the right side match substring of this compound "national [urgent] investigation" serves as "national investigation and investigation." On the other hand, binary relation sort information 3b has binary relation (a price is investigated). the binary relation (a price --) which binary relation sort information 3b has with the procedure same here as the gestalt of the 1st operation If the word "investigation" which constitutes investigation is made into the word "A" and the word "national [urgent] investigation" which constitutes the binary relation (national [urgent] investigation of price **) which subtree (national [- urgent] investigation of - price - of vegetable -) has is made into the word "B" "Investigation" which is the word "A" will correspond to the broader term which carried out compound expansion of the "national [urgent] investigation" which is the word "B", and will suit the conditions (b) of (4) shown in the gestalt of the 1st operation. although the new subtree (- investigation of - price - of vegetable -) which transposed by this "national [urgent] investigation" which is the word "B" to "investigation" which is the word "A" is created -- relation -- being equivalent (it being - investigation about - price [of vegetable -] -) -- since it is, frequency is added.
[0037] Moreover, when the thesaurus dictionary 10 is referred to, the broader terms of the word "vegetables" which subtree (national [- urgent] investigation of - price - of vegetable -) has are "agricultural products" and "agricultural products." Since binary relation sort information 3b has binary relation (price of agricultural-products **), the new subtree (national [- urgent] investigation of - price - of agricultural-products -) which transposed the word "vegetables" of subtree (national [- urgent] investigation of - price - of vegetable -) to the word "agricultural products" is created.
[0038] furthermore -- although the subtree (- investigation of - price - of agricultural-products -) which changed "vegetables" into "agricultural products" and changed "national [urgent] investigation" into "investigation" in subtree (national [- urgent] investigation of - price - of vegetable -) is created -- relation -- being equivalent (it being - investigation about - price [of agricultural-products -] -) -- for a certain reason, frequency is already added.
[0039] Moreover, since conditions are not suited, subtree (it is - investigation about - price [of agricultural-products -] -) does not perform conversion. Although the subtree (it is - investigation about - price [of agricultural-products -] -) which changed "vegetables" into "agricultural products" is created in subtree (it is - investigation about - price [of vegetable -] -) since (the price of agricultural-products **) is contained in binary relation sort information 3b, for a certain reason, frequency is already added.
[0040] As mentioned above, 3d (1, national [- urgent] investigation of - price - of agricultural-products -: - investigation: - price [of agricultural-products -] - 3, national [- urgent] investigation of - price - of vegetable - : 1, - investigation of - price - of vegetable - : 2) of conversion subtree information is created, and 3d of created conversion subtree information is sent to 1g of subtree selection sections.

h g cg b eb cg e e

[0041] 1g of subtree selection sections which received 3d of conversion subtree information will choose important subtree with a score, if 3d (1, national [- urgent] investigation of - price - of agricultural-products -: - investigation: - price [of agricultural-products -] - 3, national [- urgent] investigation of - price - of vegetable - : 1, - investigation of - price - of vegetable - : 2) of conversion subtree information is acquired.
[0042] Here, use the frequency of subtree as a score fundamentally and, in the case of the subtree abstracted by compound expansion and synonym expansion, let what multiplied the frequency by the score lapse rate (0-1) be a score. The score of each subtree at the time of setting a score lapse rate as 0.7 is shown in drawing 6 .
[0043] As mentioned above, the subtree (it is - investigation about - price [of agricultural-products -] -) of the subtree number 1 The subtree (it is - investigation about - price [of agricultural-products -] -) (subtree A) itself created by subtree creation section 1c The "vegetables" of the subtree (national [- urgent] investigation of - price - of vegetable -) created by subtree creation section 1c to "agricultural products" It is constituted by the thing (subtree B) which changed "national [urgent] investigation" into "investigation", and the thing (subtree C) which changed into "agricultural products" the "vegetables" of the subtree (it is - investigation about - price [of vegetable -] -) created by subtree creation section 1c. For the reason, the score of the subtree of the subtree number 1 brings the following results.
[0044]
The score of the subtree of the subtree number 1 = since the subtree (national [- urgent] investigation of - price - of vegetable -) of the frequency (1) x0.7=1+0.49+0.7=2.19 and the subtree number 2 of the frequency (1) x0.7x0.7+ subtree C of the frequency (1)+ subtree B of Subtree A is not what was changed, the score is set to 1 which is the value of frequency.
[0045] Furthermore, in order that the subtree (national [- urgent] investigation of - price - of agricultural-products -) of the subtree number 3 may change into "agricultural products" the "vegetables" of the subtree (national [- urgent] investigation of - price - of vegetable -) created by subtree creation section 1c, the score of the subtree of the subtree number 3 is set to 0.7 which multiplied frequency by 0.7.
[0046] Moreover, the subtree (it is - investigation about - price [of vegetable -] -) of the subtree number 4 The subtree (it is - investigation about - price [of vegetable -] -) itself created by subtree creation section 1c And since it is constituted by what changed "national [urgent] investigation" of subtree (national [- urgent] investigation of - price - of vegetable -) into "investigation", the score of the subtree (it is - investigation about - price [of vegetable -] -) of the subtree number 4 is set to 1+1x0.7=1.7.
[0047] As mentioned above, the score of the subtree (it is - investigation about - price [of agricultural-products -] -) of the subtree number 1 is the most expensive, and this subtree is chosen as important subtree. The selected important subtree (it is - investigation about - price [of agricultural-products -] -) is sent to 1h of phrase composition sections.
[0048] 1h of phrase composition sections which received important subtree (it is - investigation about - price [of agricultural-products -] -) creates epitome 3e (the prices of agricultural products are investigated) from the received important subtree. Thus, also in this gestalt, the same effectiveness as the gestalt of the 1st operation can be acquired.
[0049] In addition, the above-mentioned processing facility is realizable by computer. In that case, the contents of processing of the function which document epitome equipment should have are described to the program recorded on the record medium which can be read by computer. And the above-mentioned processing is realized by the computer by executing this program by computer. As a record medium which can be read, there are a magnetic recording medium, semiconductor memory, etc. by computer. When circulating a commercial scene, store a program in portable mold record media, such as CD-ROM (Compact Disk Read Only Memory) and a floppy disk, and they are circulated, or it stores in the storage of the computer connected through the network, and can also transmit to other computers through a network. In case it performs by computer, the program is stored in the hard disk drive unit in a computer etc., and it loads to main memory and performs.
[0050]

[Effect of the Invention] As explained above, in this invention, a related analysis means extracts binary relation from the text of two or more sentence document. Binary relation is counted with a related count means, a parse tree is created with a parse tree creation means, and subtree is created with a subtree creation means. With a subtree conversion means It changes into the subtree relevant to the binary relation which counted subtree with the related count means, and new subtree is generated. With a subtree selection means Subtree to which the score was given to subtree and the high score was given is made into important subtree. For a phrase composition means delivery and a phrase composition section means In order to compound an epitome from the sent important subtree, the subtree created can be narrowed down only to an important thing and it becomes possible as a result to generate an epitome phrase at a high speed by small capacity.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] In the document epitome equipment which creates the epitome of two or more inputted sentence document, and performs the output A related analysis means to perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and to extract the binary relation of words, A related count means to count the frequency of said binary relation which said related analysis means extracted, A parse tree creation means to create a parse tree from said binary relation which said related analysis means extracted, A subtree creation means to create subtree from said parse tree which said parse tree creation means created, The subtree conversion means which changes into the subtree relevant to said binary relation among which said related count means counted said subtree which said subtree creation means created, and creates new subtree, Document epitome equipment characterized by having a subtree selection means by which give a score to said subtree and said score chooses said high subtree as important subtree, and a phrase composition means to compound an epitome by said important subtree which said subtree selection means chose.
[Claim 2] The binary relation which transposed the word of said binary relation which said subtree which said subtree creation means created has to the synonym said subtree conversion means When in agreement with said binary relation counted with said related count means, Said new subtree is created by transposing the word of said binary relation which said subtree which said subtree creation means created has to said synonym. The binary relation which transposed the word of said binary relation which said subtree which said subtree creation means created has to the broader term When in agreement with said binary relation counted with said related count means, Said new subtree is created by transposing the word of said binary relation which said subtree which said subtree creation means created has to said broader term. The binary relation which transposed the word of said binary relation which said subtree which said subtree creation means created has to twin language Document epitome equipment according to claim 1 characterized by creating said new subtree by transposing the word of said binary relation which said subtree which said subtree creation means created has to the word of one high order when in agreement with said binary relation counted with said related count means.
[Claim 3] It is document epitome equipment according to claim 1 which has further the word relation record means which recorded the high order low order relation of a word, and is characterized by said subtree conversion means changing said subtree according to the contents of record of said word relation record means.
[Claim 4] Said subtree conversion means is document epitome equipment according to claim 1 characterized by changing said subtree by carrying out compound expansion of said subtree.
[Claim 5] Said subtree selection means is document epitome equipment according to claim 1 characterized by using frequency as said score in not being that from which said subtree was changed, and using as said score the value which multiplied said frequency by the score lapse rate when said subtree is changed.
[Claim 6] In the document epitome approach of creating the epitome of two or more inputted sentence

document, and performing the output Perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and the binary relation of words is extracted. Count the frequency of said binary relation and a parse tree is created from said binary relation. Create subtree from said parse tree, and change into the subtree relevant to said binary relation which counted said subtree, and new subtree is created. The document epitome approach characterized by giving a score to said subtree, and for said score choosing said high subtree as important subtree, and compounding an epitome by said important subtree.

[Claim 7] Perform the dependency analysis of said text which extracted the text and was extracted from two or more inputted sentence document, and the binary relation of words is extracted. Count the frequency of said binary relation and a parse tree is created from said binary relation. Create subtree from said parse tree, and change into the subtree relevant to said binary relation which counted said subtree, and new subtree is created. The record medium which recorded the program to which the function which a score is given to said subtree, and said score chooses said high subtree as important subtree, and compounds an epitome by said important subtree is made to carry out to a computer.

[Translation done.]